Chapter Four- Preliminary Data Analysis and Discussion

4.1 Introduction

The purpose of this chapter is to present the results of some statistical approach and discuss them very briefly. The chapter consists of six sections including this introduction and proceeds as follows. First, in Section 4.2 the results of the pilot study are presented. In section 4.3 stages of data analysis are graphically presented. Section 4.4 presents the results of the pre-analysis stage which include data screening and cleaning, assessing normality, and analysis of reliability. In section 4.5 the basic company data of the sampled firms are presented and described for the purpose of providing background information for better understanding of the analysis of the following chapters. Section 4.6 closes this chapter with a conclusion.

4.2 Pilot Study

In order to ensure that the instructions, questions, and scale items are clear, the survey measurement was pilot-tested on the same type of people who would be used as respondents in the main study (e.g. CEOs, Senior Managers, Operations Managers, etc.). Although, the measurement instrument was adopted from the recent study conducted by Terziovski and Samson in 2007 it was necessary to ensure that the respondents of this study can understand the questionnaire items, and respond appropriately. Therefore, pilot-test was done to pick up any questions or items that may offend potential respondents and to identify anything that might go wrong all along the process of data collection. For this reason, the questionnaire was administered to 20 respondents (with time interval) and upon completion of the

questionnaire they were asked to express their opinions on the instrument. The following findings have been discovered through pilot study:

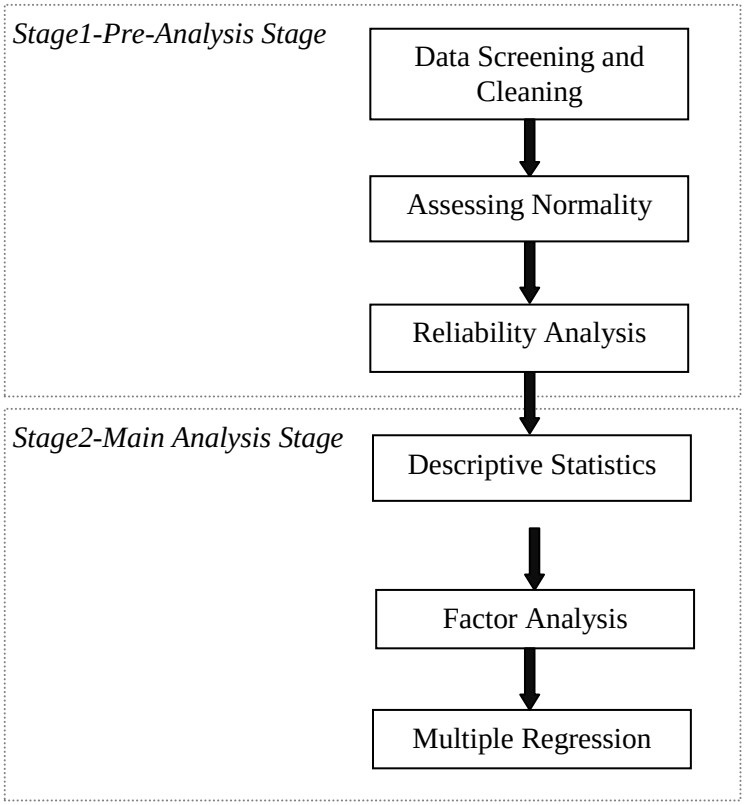
a. There were no requests for clarifications from the respondents. Therefore, it was accepted that each item on the questionnaire were understandable to the respondents.
b. The average time for each respondent to complete the questionnaire was between 20 to 25 minutes. They complained about the length of the instrument and mentioned that "CEOs do not have twenty minutes to devote to this matter!" Therefore, I had to remove three constructs comprising 37 items from the instrument. These items belonged to constructs: New Product Development, E-Commerce, and Sustainable Development Orientation which were all grouped under heading Enablers.
c. Statistical testing was done to assure that the instrument had reliability. Reliability analysis was carried out and Cronbach's alpha coefficients of 0.7 and above were found for all measured items and deemed acceptable.

Consequently, the measurement instrument comprises 2 Main constructs: Innovation Capability as the Independent Construct and Innovation Performance as the Dependent Construct. The independent construct comprises eight sub-constructs; namely, Leadership and Strategy, Employee Competence, Information and Organization Intelligence, Culture and Climate, Organization Structure, Processes, Systems and Standards, Market and Customer Orientation, Creativity and Idea Management, and finally Management of Technology. Overall, the construct-Innovation Capability includes 100 items which together affect the dependent construct- Innovation Performance which comprises eight items.

4.3 Stages of Data Analysis

The process and steps taken to analyze the quantitative data are clearly demonstrated in Figure 4.1 below.

Figure 4.1: Stages of Data Analysis



As Figure 4.1 depicts, there are two main stages of data analysis in this study. Stage One attempts to do preliminary statistical analyses such as data screening and cleaning, assessing normality and establishing the preliminary reliability, while Stage Two aims at conducting more advanced statistics through FA and MR to derive the innovation factors and build the Best Innovation Model (BIM).

4.4 Stage One-Pre-Analysis Stage

4.4.1 Screening and Cleaning the Data

Before analyzing the data it is essential to check the data and set for errors. This is the very first stage known as pre-analysis stage and it involves screening and cleaning the data comprising three basic levels: 1) checking for errors; 2) finding the errors in the data file; and 3) Correcting the error in the data file.

All three steps have been taken for all variables (both categorical and continuous) employing Frequencies procedure. For example, the item: 'Respond quickly to customer needs', which is under Market and Customer Orientation Variable, the maximum value from Descriptive Statistics table generated by SPSS 17 was '43' which was way above the maximum value (5) defined in the codebook. Therefore, through the Edit function the case was identified (case number 17) and the correct value (3) was entered.

Furthermore, since missing observations can be problematic, and to avoid this problem, most of the missing values have been replaced with estimates computed using "mean distribution method" as recommended by Coakes and Steed (2007, p.44), therefore, generating a clean, error-free data set.

4.4.1.1 Reducing the Influence of Outliers

At this stage, the outliers were identified on each individual item and on each construct as the sum of its items. There are several strategies for reducing the impact of outliers recommended by statisticians. Yet, before resorting to these strategies, it is highly recommended to first check the data for the case to make sure that they were accurately entered into the data file (SPSS 17). For univariate outliers, our option for reducing the

impact was to change the score(s) on the variable(s) for the outlying cases so that they were deviant, but not as deviant as they were. For instance, we assigned the outlying case a raw score on the offending variable that was one unit larger or smaller than the next most extreme score in the distribution (Tabachinick & Fidell, 2007). For continuous variables, we assigned the mean or average of the group to which the outlier belongs. Doing so, it helped almost Ten (10) cases to survive in the sample without threatening the reliability and precision of SPSS statistical procedures.

4.4.2 Assessing Normality

The assumption of normality, normal distribution, is a prerequisite for many inferential statistical techniques. Normal is used to describe a symmetrical, bell shaped curve, which has the greatest frequency of scores in the middle, with smaller frequencies towards the extremes (see Gravetter & Wallnau, 2000, p.52). According to Coakes and Steed (2007, p. 31) there are a number of different ways to explore this assumption graphically. These include: Histograms, Stem-and-leaf plots, Boxplots, Normal probability plots, Detrended normal plots. Moreover, a number of statistics are available to test normality. These are mainly Kolmogorov-Smirnov statistic, with Lilliefors significance level and the Shapiro-Wilk statistic, Skewness, and Kurtosis.

However, mentioned by several statisticians (see Hair et al., 1998; Coakes & Steed, 2007; Tabachnick & Fidell, 2007) normality can be assessed to some extent by obtaining skewness and kurtosis values. Therefore, in this study the main focus to determine normality of the data has been mainly on the statistics of Skewness and Kurtosis as well as Histograms and Boxplots; yet, other procedures have been employed as well. Consequently, the test of normality was run for every individual item on the questionnaire first, and later, was

repeated for each construct, of the study individually. Table 4.1below shows the results of Skewness and Kurtosis values which have been taken as the major determinant of the normal distribution. In fact, skewness value provides an indication of the *symmetry* of the distribution, while Kurtosis provides information about the *peakedness* of the distribution. Hair et al. (1998) asserts that the distribution is perfectly normal if both values for skewness and kurtosis are *zero* which is rather an uncommon occurrence in the social sciences (see Hair et al., 1998; Tabachnick & Fidell, 2007, p. 79). Therefore, a measure of skewness of ± 1 is usually regarded as a strong deviation from normality. As demonstrated by Tables 4.1 for total variables (constructs which are formed by summing their items which operationalise them) of the study, it is evident that almost all variables are skewed that is there is no perfect distribution observed according to the results. In this study, both values of skewness and kurtosis swing between -1 and +1, and most of the values are close to zero; hence, normal distribution can be assumed and parametric tests can be used to analyze the data. This assumption is also supported by Pallant (2005, p. 58): "Many scales and measures used in the social sciences have scores that are skewed, either positively or negatively. This does not necessarily indicate a problem with the scale, but rather reflects the underlying nature of the construct [in this case: innovation capability] being measured. Life satisfaction measures, for example, are often negatively skewed, with most people being reasonably happy with their lot in life." Table 4.1 demonstrates the results.

Table 4.1: Statistics Results for Normality

Variable No.	Mean	SD	Skewness	Kurtosis
1	36.83	5.147	-.231	.213
2	22.40	2.26	-.087	.260
3	93.71	16.43	-.435	.207
4	48.88	7.38	-.378	.149
5	51.85	9.53	-.223	.250
6	57.52	8.44	-.153	.654
7	37.10	5.68	-182	.404
8	18.29	3.10	-.378	-.457
9	23.63	5.36	.069	-.784

Note: 1) Leadership and Strategy, 2) Employee Competence, 3) Info. and Organization Intelligence, 4) Culture and Climate, 5) Organization Structure, Processes, Systems, and Standards, 6)Market and Customer Orientation, 7) Creativity and Idea Management, 8) Management of Technology, 9) Innovation Performance Measurement

Furthermore, Normal Q-Q Plot of each variable shows that the cases fall more or less in a straight line. This is in congruence with Coakes and Steed (2007) who claim that "in a normal probability plot, each observed value is paired with its expected value from the normal distribution; [therefore] if the sample is from a normal distribution, then the cases fall more or less in a straight line (p.34)." For the SPSS results please see Appendix 1.

4.4.3 Reliability Analysis

There are several different reliability coefficients. One of the most commonly used ones is Cronbach's alpha, which is based on the average correlation of items within the test if the items are standardized. If the items are not standardized, it is based on the average covariance coefficient and it ranges from 0 to 1.

In this study, SPSS 17 output produced Cronbach's alpha of 0.98 (N=85) for all IVs (100 items); and 0.98 (N=85) for all IVs and DVs, i.e. 108 items, altogether as a set. Although this Cronbach's alpha demonstrates a very high level of reliability, we cannot stop

here. Therefore, in order to make sure that this reliability figure is robust and healthy the reliability test on the items under each construct was run. This allows the researcher to look at each item's reliability, and decides which items to remove to improve the reliability of the construct as a whole. As a result of this, overall, four items were removed. These items are presented in Table 4.2 which demonstrates the detailed results of reliability analysis.

Table 4.2: Detailed Results on Reliability Analysis, (N=85, 108 Items)

No.	Variables Deleted	Related Construct	Constructs' Reliability before Deletion	Constructs' Reliability after Deletion
1	We adopt a deliberate (top down) strategy	Leadership	0.87	0.90
2	We rely principally on experience-based intuition when making major operating and strategic decisions	Culture and Climate	0.90	0.91
3	We offer workers legal rights in Intellectual Property, IP, they create	Creativity and Idea Management	0.90	0.92
4	We rely on 'off-the-shelf' technology for our competitive needs	Management of Technology	0.79	0.87

In addition to the above details on reliability analysis, Table 4.3 below demonstrates the results of overall reliability on each construct including the number of items kept under each.

Table 4.3: Reliability Analysis, (N=85, 104 Items)

No.	Constructs	Type	Number of Items	Cronbach's Alpha
1	Leadership	IV	9	0.90
2	Employee Competence	IV	6	0.88
3	Info. and Org. Intelligence	IV	26	0.94
4	Culture and Climate	IV	12	0.91
5	Org. Structure, Processes, Systems, and Standards	IV	15	0.91
6	Market and Customer Orientation	IV	15	0.93
7	Creativity and Idea Management	IV	9	0.92
8	Management of Technology	IV	4	0.87
9	Innovation Performance	DV	8	0.73

According to Pallant (2005, p. 90), "ideally, the Cronbach's alpha coefficient of a scale should be above 0.7". As table 4.3 shows, the construct "Information and Organizational Intelligence" has the highest Cronbach's alpha, as high as 0.94, and the Dependent Construct "Innovation Performance" has the lowest Cronbach's alpha, as low as 0.73. Therefore, as the minimum reliability of the constructs is above 0.7, it can be assumed that all constructs of the instrument produced an acceptable reliability level.

Moreover, after removing four items (variables) from the ICS (measurement instrument), the overall reliability of the test for 104 items has been established at 0.98 which is also an acceptable, very strong reliability.

4.5 Basic Company Data

4.5.1 Demographic Characteristics

Demographic statistics describes the characteristics of the sample which are elaborated on in the Methodology section. Table 4.4 presents the detailed results of demographic data.

As shown in Table 4.4, the majority of the respondents are male (66 per cent). Most of the respondents' age fall above 36-years old (82.3 per cent). Also, the ethnicity of the majority of the respondents is Chinese (58.8 per cent). This is not surprising as the economy of Malaysia is mainly dominated by Chinese-owned businesses while the politics is dominated by the Malay ethnic group. Lee (2000, p. 17) states that "awareness of the limited economic pie and impelled by the consociational politics led the political leaders to adopt a gradualist strategy to increase Malay economic participation".

Most of the sampled firms are private businesses (38.8 per cent). Moreover, most of the firms have been established for about 20 years ago (29.4 per cent). However, many of them have the age which ranges from 3 to 5 years (18.8 per cent) or fall under the age band

of 11-15 (17.6 per cent). There are only 5 firms which are very young and newly established (6 per cent).

Table 4.4: Demographics of the Sampled Firms

Item	N	Frequency	Percent %
Gender	85		
Male		56	65.9
Female		29	34.1
Respondent's Age-Band	85		
<35		37	43.5
36-45		33	38.8
46-55		10	11.8
56>		5	5.9
Ethnicity	85		
Chinese		50	58.8
Indian		16	18.8
Malay		15	17.6
European		1	1.2
Others		3	3.5
Firm's Ownership	84		
Publicly Owned		11	12.9
Privately Owned		33	38.8
Private/Family Business		15	17.6
Government-Linked Co.		1	1.2
Foreign Ownership		12	14.1
Mixed Ownership/JV		12	14.1
Firm's Age	83		
<3		5	5.9
3-5		16	18.8
6-8		6	7.1
9-10		2	2.4
11-15		15	17.6
16-20		14	16.5
>20		25	29.4

4.5.2 Industry Representativeness

Table 4.5 shows the results of industry representativeness of the sampled firms. According to this table, the majority of the sampled firms are from service industry (38 per cent). Out of these, the majority of 13 firms which represent 15.3 per cent of the service industry are from banking sector. Also, 20 of the firms are sampled from manufacturing industry. These manufacturers represent 23 per cent of the sample.

Table 4.5: Industry Representativeness by Firms

Industry	Frequency	Percent %
Services	32	37.6
Software	3	3.5
Banking	13	15.3
Insurance	1	1.2
Trade-Export/Import	4	4.2
Telecommunication	1	1.2
Manufacturing	20	23.15
Construction	5	5.9
Others	6	7.1
Total	85	100

4.5.3 Firm Size

Firm size in this study is analyzed based on the number of employees and the total annual sales. Table 4.6 shows the results of firm size analysis based on the number of employees within Malaysia.

Table 4.6: Firm Size: Number of Employees (Malaysia)

Number of Employees	Frequency	Percent %
1-5	4	4.7
6-19	11	12.9
20-50	14	16.4
51-150	13	15.3
151-300	6	7.1
301-450	6	7.1
451-above	28	32.9
Total	82	96.5
Missing	3	3.5
Total	85	100

According to Table 4.6, large firms (32.9 per cent) having 451 employees and above represent the majority of the sampled firms. Also, only 4 firms (5 per cent) are considered as micro-enterprise[1].

[1] Based on SMECorp classification available at http://www.smidec.gov.my/node/33

Table 4.7 demonstrates the results of firm size analysis based on the number of employees from sampled firms' branches overseas.

Table 4.7: Firm Size: Number of Employees (Overseas)

Number of Employees	Frequency	Percent %
1-5	4	4.7
6-19	2	2.4
20-50	4	4.7
51-150	6	7.1
151-300	1	1.2
301-450	2	2.4
451-above	19	22.4
Total	38	44.7
Missing	47	55.3
Total	85	100

According to Table 4.7, the majority of firms (55.3 per cent) did not provide any answer to this question. This can be interpreted as they have no branches or employees overseas. Apart from this, large firms form the majority of the respondents accounting for 22.4 per cent of the valid questionnaires.

Table 4.8: Firm Size: Total Annual Sales[1] (Malaysia)

Total Annual Sales	Frequency	Percent %
<200,000	3	3.5
200,000-250,000	2	2.4
250,000-1Mil	9	10.6
1Mil.-5Mil.	12	14.1
5Mil.-10-Mil.	7	8.2
10Mil.-25Mil.	11	12.9
25Mil.-above	31	36.5
Total	75	88.2
Missing	10	11.8
Total	85	100

Note1: Amounts are in Malaysian Ringgit

Table 4.8 represents the results of firms' size based on the total annual sales within Malaysia. The majority of firms (36.5 per cent) of the firms have a total annual sale of 25 million Malaysian Ringgits and above. Therefore, the majority of the firms participating in

this study are classified as large organizations[2]. Only 3 firms (3.5 per cent) have a total annual sale of less than 200,000 Malaysian Ringgits. These firms can be classified as micro-enterprises[3] and form the minority of the sampled firms.

Table 4.9: Firm Size: Total Annual Sales[1] (Entire Firm)

Total Annual Sales	Frequency	Percent %
<200,000	1	1.2
200,000-250,000	0	0
250,000-1Mil	3	3.5
1Mil.-5Mil.	8	9.4
5Mil.-10-Mil.	4	4.7
10Mil.-25Mil.	2	2.4
25Mil.-above	33	38.8
Total	51	60
Missing	34	40
Total	85	100

Note1: Amounts are in Malaysian Ringgit

Table 4.9 shows the results of firm size based on the entire total annual sales of the sampled firms. It is evident that the majority of the firms have total annual sales of 25 million Malaysian Ringgits and above, therefore, representing large organizations. There is only one firm which has total annual sales of less than 200,000 Malaysian Ringgits which can be classified as a micro-enterprise[4]. Moreover, 40 per cent of the firms did not respond to this question as they might have considered this piece of information confidential.

[2] Based on SMECorp classification available at http://www.smidec.gov.my/node/33
[3] Based on SMECorp classification available at http://www.smidec.gov.my/node/33
[4] Based on SMECorp classification available at http://www.smidec.gov.my/node/33

4.6 Conclusion

This chapter aimed at providing the brief statistics of preliminary stages of data analysis and describing the characteristics of the sampled firms such as their ownership type, age, size, as well as respondents' age, sex, and ethnic background. Based on the results it can be concluded that the sample of this study most tends to represent large, privately-owned firms from service industry.